# Business Problem:

Mr. John is re-locating to Bangalore, India along with his family. He is looking for a suitable accommodation for rent. He has 4 preferred locations in Bangalore, which are:

* 1. Yelahanka.
  2. White Field.
  3. Banashankari.
  4. Hebbal.

He is willing to rent a house in any of these 4 locations. Provided, **all** the below conditions should be satisfied.

* As he is a Foodie, he is looking for nice **Restaurants.**
* He needs a departmental **Stores** or a Super s.
* A Metro/Bus/Train **station**.
* He needs a **Park** in his neighbourhood, for his relaxation.
* He is very much interested in workouts, and needs a full equipped **Gym**

As he is new to the city, he is not sure where to rent a house.

# Need for Data Science to solve this problem:

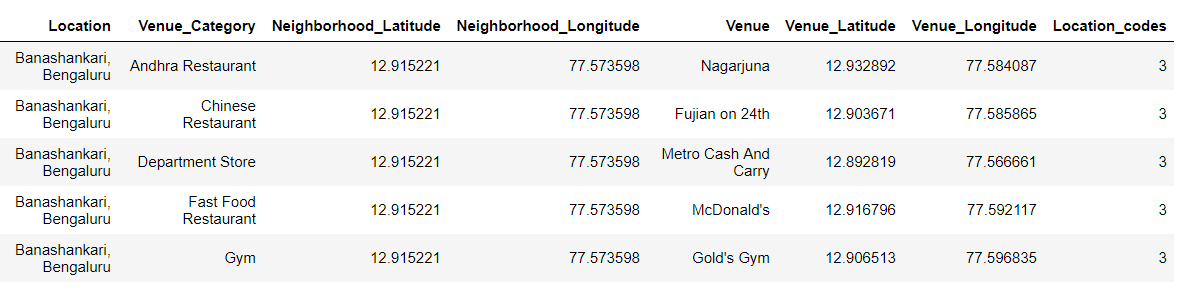
Mr. John has 4 preferred locations and has strict conditions to be satisfied to choose one of these locations. As a Data Scientist, I’ll use the available libraries and will rank the locations, based on his conditions. I’ll be able to arrive at an Best locality, which will be the suggested location to Mr. John.

# Data:

To solve this problem, I’ll use:

* Data from Foursquare API, to get the neighbourhoods of each preferred location.
* Use plotting libraries to plot each neighbourhood in the map, depicting the spread of preferences in his desired locations.
* Based on the preferences of Mr.John, I’ll check if all the conditions are satisfied for his renting.
* The best ranked location will be the one which has all the conditions met.
* This top location will be suggested for Mr. John for consideration.

# Sample data: (From Foursquare API).



# Methodology:

On a High level, the approach can be divided into following sections:

1. Data Pre-processing, which includes:
   1. Data Collection.
   2. Data Formatting.
   3. Data Labelling.
2. Feature Extraction:
   1. Extract the required features for this use case.
3. Modelling
   1. Define the business problem for a machine to understand.
4. Optimization
   1. Get the best model.
   2. Data Evaluation.

The below are the logical steps used to resolve this problem:

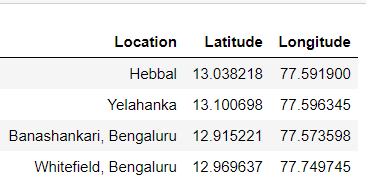
1. Import all the necessary libraries:

* Numpy
* Pandas
* Geopy, Geocoders, Nominatim
* Json, json\_normalize
* Folium
* Re
* Matplotlib, colors.

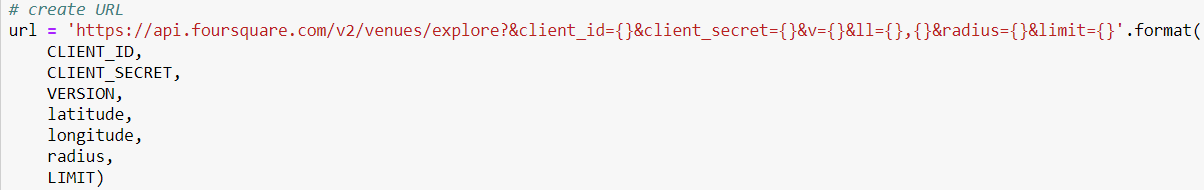
1. As a first step, collect all the locations, which Mr. John is interested to rent a house. This will be source of data for next steps.

# 

1. Using Geopy Library (geolocator, Nominatim) get the Geographical Latitude and Longitude of each location listed in Step (2), as shown below:



1. Using the Latitude and Longitudes from above step (3), hit the Foursquare API URL, to get the neighbourhood Venues of each of the above locations.

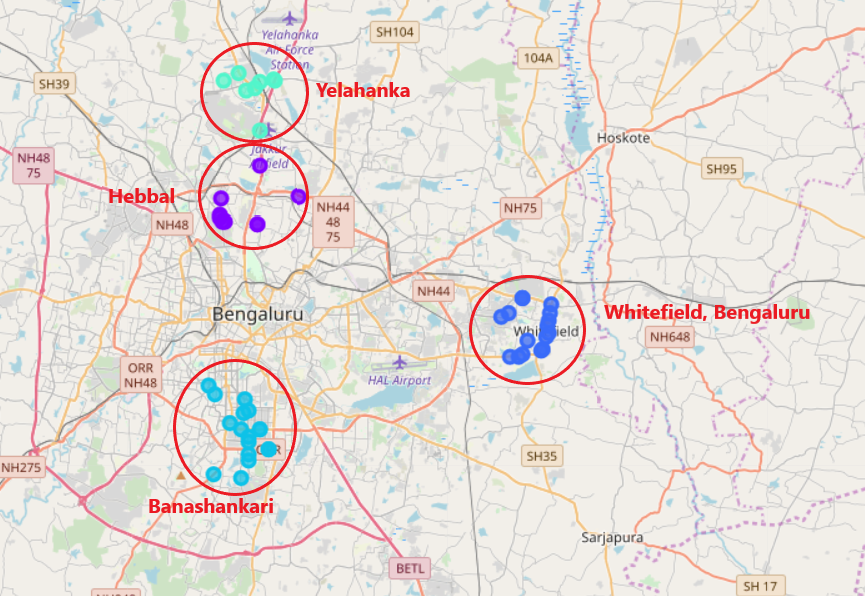


1. Get all the nearby Venues for each location into a Dataframe, which will be source of data for next steps.
2. Filter the Venue Categories based on the preferences of the user.

*User Preferences of Neighbourhood: Should have Restaurant, Gym, Park, Stores and Metro/Bus Station*



1. Pre-process the string “Locations” to numeric variables, so that they can be used as Markers in the map.
2. Plot the map to show the Venues in each location.
3. Based on the nature of the venue, Categorize the venues using Python function.
4. Cluster each locality, along with the preferences using Folium

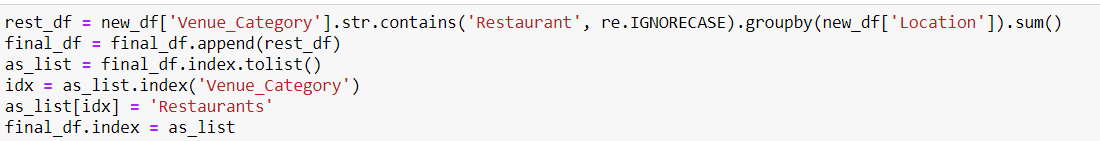


1. Out of all the Venue Categories, group similar categories and get the count of each category.

Example:

* Indian Restaurant **or** South India Restaurant **or** Hyderabadi Restaurant, all fall under and the same category of “Restaurants”.
* Gym or Fitness Centre will be categorized as “Gym/Fitness Centre”.

Create a final dataframe “final\_df” using such grouped data.



# 

# Results:

From the above final data set, we see that “Hebbal” is the location, which satisfies **all** the preferences set by Mr. John. Hence the location: “Hebbal” will be suggested to Mr. John, to rent a house.

Locations: “Banashankari” and “Yelahanka” will be ranked at 2nd preferences, as they have some missing expectations .

Finally, the location: “Whitefield” will be the last choice as it lacks nearbyparks and Bus/Railway/Metro stations.

# Conclusion:

This report uses excellent Graphical and methodological analysis to arrive at the right conclusion.